6. (Take-home, 30 points) For this problem please upload a pdf file with your solution to Canvas by Monday February 6th at noon.

   Two univariate time series y_t and z_t follow the *coupled* models over time $t = 1, ..., T$ given by:

   $$y_t = \phi y_{t-1} + \gamma z_t + \nu_t, \quad \nu_t \overset{i.i.d.}{\sim} N(0, v) \tag{1}$$

   and

   $$z_t = \theta z_{t-1} + \epsilon_t, \quad \epsilon_t \overset{i.i.d.}{\sim} N(0, w). \tag{2}$$

   Equivalently, we have $\Phi(B)y_t = \gamma z_t + \nu_t$, and $\Theta(B)z_t = \epsilon_t$ where $\Phi(B) = 1 - \phi B$, and $\Theta(B) = 1 - \theta B$ and B is the backshift operator. The two AR coefficients have values such that $|\phi| < 1$ and $|\theta| < 1$.

   (a) (4 points) The model equations above imply a *marginal* model for the y_t series alone when z_t is not observed. Show that this marginal is given by

   $$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \eta_t, \tag{3}$$

   where $\alpha_1 = \phi + \theta$ and $\alpha_2 = -\phi\theta$, and η_t is a normal zero-mean random quantity. Comment on the values of the α_1, α_2 coefficients in cases when ϕ, θ are positive and <u>both</u> close to 1. For this case, what kind of behavior you would expect to see in the y_t series?

   (b) (9 points) The model in (3) would be an AR(2) if the η_ts where independent over time, but this is not the case. Note, however, that we can approximate the structure of the y_t series when we do not observe z_t using an AR process. Consider the series $y_{1:T}$, for $T = 400$, available on Canvas. Fit a Bayesian AR(4) to this series using the conditional likelihood and a reference prior. Summarize your posterior distribution on the following quantities: (i) the AR coefficients (ii) the observational variance for this AR(4) model (iii) the posterior for the two reciprocal roots with the largest moduli and (iv) the posterior for the spectral density of y_t based on this AR(4) representation. In addition, provide the posterior mean and 95% credible intervals of the 5-steps-ahead forecast of y_t based on this AR(4) analysis. Finally, what estimated values of ϕ and θ could you derive based on this AR(4) approximations (point estimates or interval estimates are fine).

   (c) (6 points) Now use the series $z_{1:T}$, $T = 400$, available on Canvas and fit a model of the form (1). For this model assume that ϕ, γ and v are unknown. Obtain the posterior distributions for these model parameters using a likelihood function conditional on the first observation y_1 and a reference prior. Summarize the posterior distribution for (i) ϕ; (ii) γ and (iii) v. Also, assuming that you have $z_{401} = 0.106, z_{402} = 1.879, z_{403} = 1.56, z_{404} = 2.07$ and $z_{405} = 0.66$, provide a summary of the 5-steps-ahead forecast of y_t based on this model.

(d) (9 points) Finally, consider fitting a NDLM model to the y_t series, where the model is given by $\{\mathbf{F}, \mathbf{G}, v, \mathbf{W}_t\}$, such that $\mathbf{F} = (1, 0)'$,

$$\mathbf{G} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

Assume a normal-inverse Gamma prior of the form $(\boldsymbol{\theta}_0 | v, \mathcal{D}_0) \sim N(\mathbf{m}_0, v\mathbf{C}_0^*)$, and $(v|\mathcal{D}_0) \sim IG(n_0/2, (n_0 s_0)/2)$. Specify your values of $\mathbf{m}_0$, $\mathbf{C}_0^*$, n_0 and s_0.

i. Specify $\mathbf{W}_t$ using a discount factor $\delta \in (0, 1]$. Justify your choice of δ.
ii. Summarize the distribution of $(v|\mathcal{D}_{400})$.
iii. Summarize the filtering distribution $(\theta_{t,1}|\mathcal{D}_t)$ for $t = 1 : 400$ providing the mean and 95% credible intervals at each time t.
iv. Summarize the smoothing distribution $(\theta_{t,1}|\mathcal{D}_{400})$ for $t = 1 : 400$ providing the mean and 95% credible intervals at each time t.
v. Summarize the filtering distribution $(\theta_{t,2}|\mathcal{D}_t)$ for $t = 1 : 400$ providing the mean and 95% credible intervals at each time t.
vi. Summarize the smoothing distribution $(\theta_{t,2}|\mathcal{D}_{400})$ for $t = 1 : 400$ providing the mean and 95% credible intervals at each time t.
vii. Summarize the the 5-steps-ahead forecast distribution for $(y_{400+k}|\mathcal{D}_{400})$, $k = 1 : 5$, based on this model (include the mean and 95% credible intervals).

(e) (2 points) Comment on the advantages and distadvantages of the 3 modeling approaches you considered above.